BLOGGER Data

Data Set Information:

In this paper, we look for to recognize the causes of users tend to cyber space in Kohkiloye and Boyer Ahmad Province in Iran. Collecting information to form database is done by questionnaire. This questionnaire is provided as oral, written and also programming of a website which includes an internet questionnaire and the users can answer the questions as they wish. They entered their used websites, blogs and social networks during the day. After collecting questionnaires, the wed addresses are gathered to get expected results. And finally, their trustfulness is checked by analyzing their used web pages. As the results were same, for getting better and noiseless response, they will put in database

Attribute Information:

We considered the following parameters as questions: age, education, political attitudes, blog topic, and the type of the identity in internet, the influence of managersâ€™ inefficiency on tendency, the effect of inefficient media on tendency, the effects of social and political conditions on tendency and finally the effect of poverty in the province on tendency. The noisy or too detailed data in database makes us far from to get proper and suitable answers of algorithms [8]. We preprocessed the data and eliminated some non-relevant data. Finally the followings are considered as the main fields which include: education, political caprice, topics, local media turnover (LMT) and local, political and social space (LPSS). The collected data are shown in Table 1. In order to get correct answer, we classify bloggers to two groups: professional bloggers and seasonal (temporary) bloggers. Professional bloggers are those who adopt blog as an effective digital media and interested in digital writing in continuous time intervals. Seasonal (temporary) bloggers arenâ€™t professional and follow blogging in discrete time periods. In this study, we review the tendency factors considering whether these people are among professional bloggers (Pro Bloggers, PB) and then, consider the other factors according to it.

Link : https://archive.ics.uci.edu/ml/datasets/BLOGGER

SVM:“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However,it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.

KNN:The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.

Perceptron:The Perceptron is a linear machine learning algorithm for binary classification tasks.

It may be considered one of the first and one of the simplest types of artificial neural networks. It is definitely not “deep” learning but is an important building block.
Like logistic regression, it can quickly learn a linear separation in feature space for two-class classification tasks, although unlike logistic regression, it learns using the stochastic gradient descent optimization algorithm and does not predict calibrated probabilities.

Random Forest Classifier

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model’s prediction